STINFO COPY

AFRL-HE-WP-TR-2005-0172

AIR FORCE RESEARCH LABORATORY

Development and Validation of the Test of Basic Aviation Skills (TBAS)

Thomas R. Carretta

Human Effectiveness Directorate
Wright-Patterson AFB, OH 45433-7022

Interim Report for May 2005 to September 2005

November 2005

Human Effectiveness Directorate
Warfighter Interface Division
System Control Interfaces Branch
Wright-Patterson AFB OH 45433

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2005-0172

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

//Signed//

MARIS M. VIKMANIS
Chief, Warfighter Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
112005	Interim	May 2005 - September 2005

4. TITLE AND SUBTITLE
Development and Validation of the Test of Aviation Skills (TBAS)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Thomas R. Carretta

5d. PROJECT NUMBER 7184

5e. TASK NUMBER 09

5f. WORK UNIT NUMBER 72

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Wright-Patterson AFB OH 45433-7022

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-HE-WP-TR-2005-0172

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Wright-Patterson AFB OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited. Clearance AFRL/WS-05-2771, 13 Dec 05.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In 1993, the Pilot Candidate Selection Method (PCSM) was operationally implemented as an adjunct to US Air Force pilot training selection methods. PCSM combined the Air Force Officer Qualifying Test (AFOQT) Pilot composite scores from the Basic Attributes Test (BAT) and a measure of prior flying experience in a regression-weighted pilot aptitude composite. Since 1993, neither the BAT hardware nor software have been updated. As with all aptitude tests, it is desirable to update test content at regular intervals to keep it current and avoid potential problems such as test compromise. In the case of computer-based tests such as the BAT, it is also desirable to update test hardware and software to avoid problems associated with normal wear to the system (e.g., calibration of the control sticks, functioning of input devices) and to take advantage of advances in computer hardware and software. The Test of Basic Aviation Skills (TBAS) was developed as a candidates BAT replacement test in the PCSM equation. The purpose of this report is to document the TBAS development process and report results of a study of its validity and incremental validity versus measures of pilot training performance when used with other operational measures of pilot aptitude (i.e., AFOQT), prior flying experience).

15. SUBJECT TERMS
Pilot aptitude, Air Force Officer Qualifying Test, Basic Attributes Test, Pilot Candidate Selection Method, Test of Basic Aviation Skills

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UNL	23	Thomas R. Carretta
					19b. TELEPHONE NUMBER *(Include area code)* (937) 656-7014

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

TABLES

PREFACE

This report describes activities performed in support of work unit 71840972 in support of USAF aircrew selection and classification. The author thanks AETC SAS/CS for their support. 1Lt Tanya S. Garcia and SSgt Leonard M. Burkhardt were instrumental in the development of the database used in this study. Further, SSgt Burkhardt was the lead programmer in software development for the Test of Basic Aviation Skills (TBAS). 2dLt Kirk Reimer provided TBAS hardware specifications. The author also thanks Mr. Kenneth L. Schwartz and Mr. Johnny J. Weissmuller (AFPC/DPPPWT) for providing supporting materials about the tests.

THIS PAGE INTENTIONALLY LEFT BLANK

DEVELOPMENT AND VALIDATION OF THE TEST OF BASIC AVIATION SKILLS (TBAS)

In 1993, the Pilot Candidate Selection Method (PCSM) was operationally implemented as an adjunct to US Air Force pilot trainee selection methods. PCSM combines the Air Force Officer Qualifying Test (AFOQT) Pilot composite, scores from the Basic Attributes Test (BAT) and a measure of prior flight experience in a regression-weighted pilot aptitude composite. Since 1993, neither the BAT hardware nor software has been updated. The original hardware consisted of a 386-based computer, monochrome monitor, two ruggedized control sticks, and a specialized response keypad (Carretta & Ree, 1993). As with all aptitude tests, it is desirable to update test content at regular intervals. This is done to keep test content current and avoid potential problems such as test compromise. In the case of computer-based tests, it is also desirable to update hardware and software to avoid problems with normal wear to the system (e.g., calibration of control sticks, functioning of input devices) and to take advantage of improvements to computer hardware and software.

Several critical issues have been identified regarding the development of a replacement for the BAT and update to the PCSM model. These include: 1) development and validation of a BAT replacement test, 2) development of a PCSM replacement model, and 3) operational implementation of the new test battery and development of supporting documentation.

Development and Validation of a BAT Replacement Test

Initial Development and Validation

AETC/SAS developed a candidate replacement battery for the BAT called the Test of Basic Aviation Skills (TBAS). The experimental TBAS battery consisted of nine cognitive, perceptual, and psychomotor subtests. These were: 1) 3-Digit Listening (3DIG), 2) 5-Digit Listening (5DIG), 3) Airplane Tracking (ATT), 4) Horizontal Tracking (HTT), 5) Airplane Tracking & Horizontal Tracking (AHTT), 6) Airplane Tracking, Horizontal Tracking, & 3-Digit Listening (AHTT3), 7) Airplane Tracking, Horizontal Tracking, & 5-Digit, Listening (AHTT5), 8) Emergency Scenario (EST), and 9) UAV Test (UAV). Brief descriptions of the TBAS subtests are provided below.

Ree (2003b) examined the validity of the TBAS against Specialized Undergraduate Pilot Training (SUPT) T-37 final outcome (pass/fail) and T-37 Total Score (based on flying grades) in a sample of 551 students. Four of the nine TBAS subtests showed validity against pilot training performance (ATT, HTT, EST, & UAV). The validities of a regression-weighted TBAS-only model based on scores from these four subtests were R = .303 versus pass/fail T-37 training and R = .331 versus the T-37 Total score.

TBAS Incremental Validity

Ree (2003a) subsequently examined the incremental validity of TBAS versus SUPT performance when used in combination with the AFOQT in a sample of 322 pilot trainees. Incremental validity analyses focused only on the five TBAS scores that demonstrated validity in Ree (2003b) (ATT Skilled Redirects, HTT Skilled Redirects, EST Skill Level, UAV Total Time, and UAV Total Correct). The training criteria included a dichotomous SUPT T-37 pass/fail score and T-37 Total Score.

Examination of the correlation matrix showed significant correlations between several of the predictors and the SUPT performance criteria. The AFOQT Pilot composite alone was significantly related to both the SUPT pass/fail score (r = .197) and to the T-37 Total Score (r = .309). The correlations of the TBAS scores with the SUPT T-37 pass/fail ranged from .102 (EST Skill Level) to .137 (UAV Total Correct) with a mean of .113. The correlations of the TBAS scores with the SUPT T-37 Total Score ranged from .132 (EST Skill Level) to .216 (HTT Skilled Redirects) with a mean of .187.

Regression models were tested to examine the incremental validity of the TBAS scores when used in combination with the AFOQT Pilot composite. Results showed that only HTT Skilled Redirects and UAV Total Correct scores provided incremental validity versus the SUPT criteria.

Although results of these studies are encouraging, they should be viewed as preliminary. The TBAS is intended to be used as component of an updated PCSM model. Scores from the AFOQT and biographical data (e.g., previous flying experience) will serve as a baseline, as these data are available independent of TBAS. In order to be included in a revised PCSM model, the TBAS scores must demonstrate incremental validity beyond the AFOQT and flying experience.

The purpose of this study was to examine the incremental validity of TBAS scores when used in conjunction with the AFOQT composite scores and previous flying experience.

METHOD

Participants

Participants were 994 USAF pilot trainees who attended SUPT T-37 training. All had completed the AFOQT in order to apply for an officer commissioning program and for pilot training. Participants completed the TBAS after they had already been accepted into SUPT. Their TBAS scores had no effect on qualification for pilot training. The sample was predominantly male (95.0%) and White (89.5%). The T-37 graduation rate was 88.7%.

Measures

The predictors used were the AFOQT composite scores, a previous flying experience score, and scores from the TBAS subtests.

Air Force Officer Qualifying Test

The AFOQT is a paper-and-pencil multiple aptitude test battery used for officer commissioning and aircrew training qualification. Forms O through Q1/Q2 consist of 16 subtests that are combined to form five composite scores: Verbal (V), Quantitative (Q), Academic Aptitude (AA = V + Q), Pilot (P), and Navigator-Technical (N-T) (Carretta & Ree, 1996). Forms Q1/Q2 were used during this study.

AFOQT Forms S1/S2 were operationally implemented in July 2005. With their implementation, five of the 16 subtests were removed. However, Forms S1/S2 retained the factor structure and operational composites of the previous forms (Gould & Shore, 2003; Skinner & Alley, 2002).

Previous Flying Experience

A previous flying experience score contributes to the Pilot Candidate Selection Method (PCSM) score implemented in 1993 (Carretta, 2000). Flying hours are recorded using an unequal interval scale.

Test of Basic Aviation Skills

The TBAS is a computer-administered cognitive and perceptual-motor test battery designed to measure pilot aptitude. It was developed as a replacement for the Basic Attributes Test (BAT; Carretta & Ree, 1993). The TBAS battery consists of 8 subtests.

Three Digit Listening Test (3DIG). During this test, a series of numbers and letters are presented via headphones. Examinees are instructed to press a trigger on a control stick when any of three identified numbers (i.e., targets) are presented and to not respond to others (i.e., non-targets). Performance is based on response accuracy where examinees are given credit for correct responses and penalized for incorrect responses. For example, an examinee may be instructed to respond when they hear a 0, 3, or 6. If they hear "Y R Z 9 F C X 2 B 3 E 7 6 J" they should click the trigger immediately after hearing the number 3 and immediately after hearing the number 6.

Five Digit Listening Test (5DIG). This is the same as the Three Digit Listening Test except in this test the examinee is instructed to respond to five identified numbers.

Airplane Tracking Test (ATT). This compensatory tracking task measures the ability to track a moving target in two dimensions (horizontal and vertical). The image of an airplane and crosshairs appear on the computer screen. The examinee's task is to keep the crosshairs centered on the airplane. The difficulty of the task varies. Examinees are scored on how accurately they track the airplane.

Horizontal Tracking Test (HTT). This compensatory tracking task measures the ability to track a moving target on a horizontal axis. The image of an airplane and a box appear on the computer screen. The airplane moves left and right across the screen at various speeds that the examinee cannot control. Examinees are instructed to use rudder pedals to keep the airplane inside the box for as long as possible. Performance is based on how accurately examinees track the airplane.

Airplane Tracking and Horizontal Tracking Test (AHTT). This test requires examinees to simultaneously perform the ATT and HTT tracking tasks. Examinees manipulate the control stick to target an airplane moving in two dimensions while simultaneously manipulating the rudder pedals to target an airplane moving along a horizontal axis. Performance is based on how accurately both airplanes (targets) are tracked.

Airplane Tracking, Horizontal Tracking, and Three Digit Listening Test (AHTT3). In this test, examinees are required to simultaneously perform the ATT, HTT, and 3DIG tasks. The

control stick is used to target an airplane moving in two dimensions as in ATT, the rudder pedals to target an airplane moving along a horizontal axis as in HTT, and the trigger on the control stick to respond to the 3-Digit Listening (3DIG) task. Performance is based on tracking accuracy.

Airplane Tracking, Horizontal Tracking, and Five Digit Listening Test (AHTT5). This test is the same as AHTT3 except the examinee is now listening for and responding to five digits (5DIG) rather than three.

Emergency Scenario Test (EST). In this test, examinees simultaneously perform the Airplane Tracking Test and Horizontal Tracking Test and must respond to audio warnings indicating an emergency situation. Examinees are required to make certain responses on the keypad to resolve the emergency situations while continuing to perform the tracking tasks. Performance is based on tracking performance and response speed and accuracy to the emergency situations.

UAV Test (UAV). An airplane is shown flying on the computer screen with its direction indicated and a map of the ground view. Examinees are asked to identify map locations. For example, the examinee may be told the airplane is flying NE and to identify the south parking lot. Performance is based on speed and accuracy of response.

SUPT Performance

Two SUPT performance criteria were examined. The first was a dichotomous T-37 pass/fail score, scored 1 for graduates and 0 for eliminees. The second criterion was T-37 Final Score which was a weighted composite of T-37 daily flying average, check flight average, and academic average.

TBAS Apparatus

The TBAS software was hosted on a 2.80 GHz CPU computer with 512 MB RAM and a 40 GB hard drive, CD ROM and USD port removable media storage devices, and a Microsoft Windows™ XP operating system. The monitor was a 17-inch flat panel with 0.264 pixel pitch, 1280 by 1024 pixel resolution, and a sync rate of 56 Hz by 75 Hz (vertical by horizontal). The control stick was a Thrustmaster™ Model Hotas Cougar and the rudder pedals were CH Products™ Pro Pedals. The computer hardware was housed in a wooden carrel to provide a standardized test environment and reduce distractions.

Analyses

Several analyses were performed to examine the relations between the test scores and SUPT performance. Descriptive statistics (means, standard deviations) and correlations were examined. Relations among the TBAS scores were examined to determine the utility of creating composite scores that combined scores across TBAS subtests.

The data were corrected for the effects of range restriction due to prior selection for officer commissioning and pilot training. This was done to provide a better estimate of the relations among the tests scores and training criteria.

Next, several regression models were developed to examine the predictive utility of the AFOQT composites, flying experience, and the TBAS subtests versus SUPT performance. To begin, a baseline pilot candidate selection model was developed to determine the predictive utility of currently available operational scores (AFOQT and flying experience). Subsequent regression models examined whether TBAS scores incremented the validity of this baseline model. All analyses used a .05 Type I error rate. Regressions were performed using both the observed correlations and the correlations after correction for range restriction (Lawley, 1943; Ree et al., 1994). The regressions involving the T-3-7 pass/fail criterion also were corrected for dichotomization of the criterion (Cohen, 1983).

RESULTS AND DISCUSSION

Means and Standard Deviations

Table 1 summarizes the means and standard deviations for the test scores and SUPT training criteria. As the result of prior selection for officer commissioning and pilot training, the mean AFOQT composite scores were elevated above the normative value of 50 and the standard deviations were lower than the normative value of 28.29. The AFOQT composite means ranged from 0.19 (Verbal) standard deviations to 0.81 (Pilot) standard deviations above the normative values, with an average increase of 0.41 standard deviations. The AFOQT composite variances ranged from 0.36 (Pilot) to 0.64 (Verbal) of the normative values, with a mean of 0.54.

Table 1. Means and Standard Deviations of the Test Scores and SUPT Criteria

Score	Abbrev.	Observed Mean	Observed SD	Corrected Mean	Corrected SD
1. AFOQT Pilot	AFOQT-P	72.97	17.04	50.00	28.29
2. AFOQT Nav-Tech	AFOQT-N	67.47	19.19	50.00	28.29
3. AFOQT Academic	AFOQT-A	56.59	22.25	50.00	28.29
4. AFOQT Verbal	AFOQT-V	55.49	22.64	50.00	28.29
5. AFOQT Quantitative	AFOQT-Q	56.11	22.40	50.00	28.29
6. Flying Hours Code	FLYHRS	3.61	3.40	1.75	3.68
7. 3-Digit N Correct	3DIG_NC	5.91	0.39	5.89	0.38
8. 5-Digit N Correct	5DIG_NC	9.67	1.25	9.68	1.24
9. ATT N Skilled Redirects	ATT_SR	7.88	4.09	7.08	4.17
10. HTT N Skilled Redirects	HTT_SR	15.04	4.88	14.23	4.94
11. ATT/HTT N Skilled Redirects	AHTT_SR	6.49	4.54	5.90	4.57
12. ATT/HTT 3-Digit N Skilled Redirects	AHTT3_SR	12.25	7.64	10.69	7.82
13. ATT/HTT 5-Digit N Skilled Redirects	AHTT5_SR	13.11	8.15	11.31	8.41
14. Emergency Scenario Mean RT	EST_RT	2140.39	852.17	2274.79	868.45
15. UAV Mean RT	UAV_RT	116.96	51.85	127.16	52.76
16. UAV N Correct	UAV_NC	36.25	8.84	32.59	9.79
17. ATT_SR + HTT_SR Composite	AHTT_SR2	22.92	6.91	21.31	7.10
18.ATT_SR + HTT_SR + AHTT_SR Composite	AHTT_SR3	29.41	9.98	27.22	10.23
19. ATT_SR + HTT_SR + AHTT_SR + AHTT3_SR	AHTT_SR5	54.77	23.48	49.25	24.22

+ AHTT5_SR Composite					
20. SUPT T-37 Pass/Fail	T37_PF	0.888	0.315	0.810	0.324
21. SUPT T-37 Total Score	T37_TS	41.76	19.22	36.31	19.91

N = 994

Correlations

Table 2 summarizes the correlations between the test scores and SUPT training criteria. Correlations below the diagonal are observed values; those above the diagonal were corrected for range restriction using the multivariate method (Lawley, 1943; Ree, Carretta, Earles, & Albert, 1994) and the RANGEJ software (Johnson & Ree, 1994). Upon examination of the correlations, it was decided to create three TBAS composites that combined scores across subtests (AHTT_SR2, AHTT_SR3, and AHTT_SR5).

Clearly, the observed correlations are downwardly biased due to the effects of range restriction caused by prior selection for officer commissioning and pilot training based in part on applicants' AFOQT scores. For example, the observed correlation between the AFOQT Pilot composite and the SUPT training criteria were .193 for the dichotomous T-37 pass/fail score and .217 for the T-37 Total Score. After correction for range restriction, the correlation between the AFOQT Pilot composite and T-37 training criteria were .305 and .337 respectively.

Regression Analyses

Next, a baseline pilot selection model was developed that used only AFOQT composite scores and previous flying experience. Tables 3 and 4, respectively, summarize the results of the regression analyses predicting the dichotomous T-37 pass/fail criterion and the T-37 Total Score criterion. Model 1, the baseline model, included three predictors: AFOQT-P, AFOQT-Q, and FLYEXP.

Table 2. Correlation Matrix

Score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1.	1.000	0.930	0.750	0.610	0.740	0.495	0.057	0.031	0.243	0.204	0.157	0.259	0.294	-0.222	-0.240	0.527	0.297	0.270	0.296	0.305	0.337
2.	0.817	1.000	0.840	0.620	0.900	0.331	0.077	0.041	0.233	0.206	0.158	0.260	0.294	-0.250	-0.229	0.545	0.282	0.269	0.294	0.298	0.326
3.	0.544	0.768	1.000	0.900	0.910	0.228	0.083	0.054	0.194	0.147	0.121	0.213	0.254	-0.197	-0.184	0.456	0.218	0.208	0.240	0.218	0.241
4.	0.428	0.464	0.841	1.000	0.640	0.192	0.083	0.039	0.165	0.101	0.098	0.176	0.219	-0.144	-0.146	0.355	0.169	0.163	0.198	0.135	0.185
5.	0.496	0.832	0.864	0.468	1.000	0.201	0.071	0.058	0.186	0.165	0.122	0.211	0.243	-0.215	-0.185	0.471	0.227	0.214	0.238	0.254	0.268
6.	0.330	0.048	-0.035	-0.005	-0.061	1.000	-0.011	0.001	0.097	0.085	0.028	0.062	0.067	-0.046	-0.099	0.228	0.117	0.093	0.082	0.182	0.238
7.	0.043	0.083	0.086	0.080	0.064	-0.036	1.000	0.186	0.025	0.055	0.032	0.027	0.050	-0.019	-0.035	0.028	0.054	0.051	0.047	0.046	0.004
8.	0.018	0.039	0.055	0.031	0.060	-0.012	0.186	1.000	0.040	0.074	0.058	0.070	0.079	0.059	0.004	0.059	0.076	0.079	0.083	0.031	0.047
9.	0.139	0.113	0.075	0.071	0.064	0.028	0.015	0.035	1.000	0.205	0.660	0.681	0.689	-0.486	-0.105	0.232	0.732	0.804	0.798	0.123	0.142
10.	0.105	0.088	0.028	0.007	0.051	0.031	0.044	0.069	0.178	1.000	0.174	0.178	0.218	-0.170	-0.085	0.207	0.817	0.645	0.405	0.129	0.096
11.	0.084	0.073	0.034	0.032	0.034	-0.017	0.024	0.054	0.653	0.156	1.000	0.769	0.760	-0.484	-0.099	0.166	0.509	0.802	0.850	0.089	0.094
12.	0.150	0.142	0.094	0.080	0.087	-0.017	0.017	0.066	0.667	0.147	0.765	1.000	0.850	-0.544	-0.107	0.213	0.526	0.709	0.917	0.144	0.143
13.	0.163	0.144	0.105	0.101	0.091	-0.020	0.037	0.073	0.674	0.185	0.758	0.842	1.000	-0.589	-0.116	0.248	0.557	0.727	0.928	0.166	0.152
14.	-0.124	-0.158	-0.100	-0.061	-0.116	0.023	-0.011	-0.055	-0.466	-0.141	-0.471	-0.525	-0.569	1.000	0.043	-0.238	-0.405	-0.497	-0.589	-0.127	-0.118
15.	-0.153	-0.141	-0.098	-0.071	-0.092	-0.024	-0.031	0.000	-0.069	-0.051	-0.075	-0.069	-0.071	0.008	1.000	0.149	-0.121	-0.129	-0.128	-0.074	-0.093
16.	0.341	0.369	0.273	0.188	0.280	0.075	0.014	0.055	0.162	0.143	0.119	0.136	0.157	-0.176	0.258	1.000	0.281	0.271	0.267	0.242	0.246
17.	0.157	0.129	0.064	0.047	0.074	0.039	0.040	0.070	0.719	0.812	0.497	0.500	0.530	-0.376	-0.077	0.197	1.000	0.923	0.752	0.164	0.152
18.	0.146	0.123	0.060	0.048	0.067	0.019	0.038	0.073	0.795	0.633	0.800	0.694	0.712	-0.474	-0.087	0.191	0.919	1.000	0.903	0.154	0.147
19.	0.168	0.148	0.093	0.081	0.088	-0.005	0.035	0.078	0.789	0.391	0.852	0.913	0.924	-0.570	-0.084	0.180	0.737	0.898	1.000	0.168	0.160
20.	0.193	0.189	0.112	0.034	0.148	0.093	0.040	0.028	0.078	0.088	0.058	0.097	0.112	-0.086	-0.032	0.160	0.109	0.102	0.114	1.000	0.602
21.	0.217	0.207	0.124	0.057	0.146	0.146	-0.003	0.044	0.093	0.049	0.060	0.091	0.091	-0.072	-0.037	0.154	0.090	0.089	0.099	0.576	1.000

Notes. 1. The scores are: 1) AFOQT-P, 2) AFOQT-N, 3) AFOQT-A, 4) AFOQT-V, 5) AFOQT-Q, 6) FLYHRS, 7) 3DIG_NC, 8) 5DIG_NC, 9) ATT_SR, 10) HTT_SR, 11) AHTT_SR, 12) AHTT3_SR, 13) AHTT5_SR, 14) EST_RT, 15) UAV_RT, 16) UAV_NC, 17) AHTT_SR2, 18) AHTT_SR3. 19) AHTT_SR5, 20) T37_PF, and 21) T37_TS.

2. Correlations below the diagonal are observed values. Those above the diagonal were corrected for range restriction using the multivariate method (Lawley, 1943; Ree et al., 1994).

3. Observed correlations with an absolute value ≥ .052 are significant at p = .05 (1-tailed test). Observed correlations with an absolute value ≥ .073 are significant at p = .01 (1-tailed test).

Table 3. Summary of Regression Analyses for SUPT T-37 Pass/Fail Criterion

Model	R	ΔR	R_C	ΔR_C	$R_{C'}$	$\Delta R_{C'}$
1. AFOQT-P, AFOQT-Q, & FLYEXP	0.208**		0.312		0.451	
2. Model 1 + 3DIG_NC	0.210**	0.002	0.313	0.001	0.452	0.001
3. Model 1 + 5DIG_NC	0.208**	0.001	0.313	0.001	0.452	0.001
4. Mpdel 1 + ATT_SR	0.215**	0.007	0.317	0.005	0.458	0.007
5. Model 1 + HTT_SR	0.219**	0.011*	0.320	0.008	0.462	0.011
6. Model 1 + AHTT_SR	0.213**	0.005	0.315	0.003	0.455	0.004
7. Model 1 + AHTT3_SR	0.220**	0.012*	0.320	0.008	0.462	0.011
8. Model 1 + AHTT5_SR	0.225**	0.017**	0.323	0.011	0.467	0.016
9. Model 1 + EST_RT	0.212**	0.004	0.318	0.006	0.460	0.009
10. Model 1 + UAV_RT	0.208**	0.000	0.312	0.000	0.451	0.000
11. Model 1 + UAV_NC	0.228**	0.020**	0.325	0.013	0.470	0.019
12. Model 1 + AHTT_SR2	0.223**	0.015**	0.322	0.010	0.465	0.014
13. Model 1 + AHTT_SR3	0.222**	0.014*	0.321	0.009	0.464	0.013
14. Model 1 + AHTT_SR5	0.225**	0.017**	0.323	0.011	0.467	0.016
15. Model 1 + All 13 TBAS Scores	0.253**	0.047*	0.341	0.029	0.493	0.042
16. Model 1 + Stepwise TBAS Scores[a]	0.240**	0.032**	0.333	0.021	0.481	0.030

N = 994

Notes. R values were based on the observed correlations. R_C values were based on the correlations corrected for range restriction. $R_{C'}$ values were based on the correlations corrected for range restriction and for dichotomization of the T-37 pass/fail score.

[a] The scores in Model 16 (Model 1 + Stepwise TBAS scores) were AFOQT-P, AFOQT-Q, FLYHRS, UAV_NC, and AHTT5_SR.

* p < .05, ** p < .01

As shown in Table 3, using the uncorrected (observed) data, Model 1 was significantly related to the T-37 pass/fail criterion ($R = 0.208$, $p < .01$). Using the range-restriction corrected data, the multiple R increased to .312 and increased further to .451 when the data were corrected for both range restriction and the dichotomization (Cohen, 1983) of the T-37 pass/fail criterion. Models 2 through 14 examined the incremental validity of individual TBAS scores when used in conjunction with the baseline (Model 1). Seven of the 13 TBAS scores showed some incremental validity beyond Model 1 alone. The largest increment was provided by the UAV Test Number Correct score with an increment of 0.020 (Model 1: $R = 0.208$; Model 1 + UAV_NC: $R = 0.228$) using the observed (uncorrected) data. When the TBAS scores were allowed to enter in a stepwise manner (Model 16) after entering the Model 1 scores (AFOQT-P, AFOQT-Q, and FLYHRS), only two scores showed incremental validity (AHTT5_SR and UAV_NC). The observed, range restriction corrected, and fully corrected multiple correlations for Model 16 were 0.240, 0.333, and 0.481, respectively.

Similar results were obtained for the T-37 Total Score regression analyses (Table 4). Model 1 was significantly related to the T-37 Total Score criterion ($R = 0.241$, $p < .01$). After correction for range restriction, the multiple R for Model 1 was 0.351. Six of the 13 TBAS scores showed incremental validity beyond Model 1 alone. As with the T-37 pass/fail criterion analyses, the UAV Test Number Correct score (UAV_NC) showed the largest increment (0.013) when used in conjunction with Model 1 (Model 1: $R = 0.241$; Model 1 + UAV_NC: $R = 0.254$). When the TBAS scores were allowed to enter in a stepwise manner (Model 16) after entering the Model 1 scores, only the UAV_NC score showed incremental validity. The observed and range restriction corrected multiple correlations for Model 16 were 0.254 and 0.360, respectively.

Final Revised PCSM Model

The best-fitting parsimonious model for predicting SUPT T-37 pass/fail included the AFOQT Pilot composite, AFOQT Quantitative composite, FLYHRS, TBAS UAV_NC, and TBAS AHTT5_SR (see Table 3, Model 16). At first glance, it seems that TBAS could be scaled back as test administration could be limited to the *UAV Test* (UAV) and the *Airplane Tracking, Horizontal Tracking, and Five Digit Listening Test* (AHTT5). However, it should be noted that AHTT5 builds on several previous tests that provide the opportunity to practice one or more of the AHTT5 component tasks. That is, prior to testing on the *Airplane Tracking, Horizontal*

Table 4. Summary of Regression Analyses for SUPT T-37 Total Score Criterion

Model	R	ΔR	R_C	ΔR_C
1. AFOQT-P, AFOQT-Q, & FLYEXP	0.241**		0.351	
2. Model 1 + 3DIG_NC	0.241**	0.000	0.351	0.000
3. Model 1 + 5DIG_NC	0.244**	0.003	0.353	0.002
4. Mpdel 1 + ATT_SR	0.250**	0.009*	0.357	0.006
5. Model 1 + HTT_SR	0.243**	0.002	0.352	0.001
6. Model 1 + AHTT_SR	0.246**	0.005	0.354	0.003
7. Model 1 + AHTT3_SR	0.250**	0.009*	0.357	0.006
8. Model 1 + AHTT5_SR	0.249**	0.008*	0.357	0.006
9. Model 1 + EST_RT	0.246**	0.005	0.354	0.003
10. Model 1 + UAV_RT	0.242**	0.001	0.351	0.000
11. Model 1 + UAV_NC	0.254**	0.013**	0.360	0.009
12. Model 1 + AHTT_SR2	0.248**	0.007	0.356	0.005
13. Model 1 + AHTT_SR3	0.249**	0.008*	0.357	0.006
14. Model 1 + AHTT_SR5	0.251**	0.010*	0.358	0.007
15. Model 1 + All 13 TBAS Scores	0.268**	0.027	0.365	0.014
16. Model 1 + Stepwise TBAS Scores[a]	0.254**	0.013**	0.360	0.009

N = 994

Notes. R values were based on the observed correlations. R_C values were based on the correlations corrected for range restriction.

[a] The scores in Model 16 (Model 1 + Stepwise TBAS scores) were AFOQT-P, AFOQT-Q, FLYHRS, and UAV_NC.

* p < .05, ** p < .01

Tracking, and Five Digit Listening Test (AHTT5), participants complete the *Three Digit Listening Test* (3DIG), *Five Digit Listening Test* (5DIG), *Airplane Tracking Test* (ATT), *Horizontal Tracking Test* (HTT), *Airplane Tracking and Horizontal Tracking Test* (AHTT), and the *Airplane Tracking, Horizontal Tracking, and Three Digit Listening Test* (AHTT3). It is not known how performance on AHTT5 would be affected if these tests were removed from the TBAS battery. Is the small increment in validity (0.012) afforded by the AHTT5_SR score sufficient to warrant having to administer the other tests (Model 1 + UAV_NC: R = 0.228; Model 1 + UAV_NC + AHTT5: R = 0.240)?

Two possible alternatives to Model 16 are to use only the TBAS UAV_NC score along with the baseline scores (Model 11) or to identify another TBAS score that is almost as incremental as AHTT5, but would not require administering so many of the TBAS subtests. Table 5 summarizes the results of several alternate regression models. Model 1 (new baseline) included the AFOQT-P, AFOQT-Q, FLYHRS, and UAV_NC scores. Models 2-6 examined the incremental validity gained by adding other TBAS scores to Model 1.

Table 5. Summary of Additional Regression Analyses for SUPT T-37 Pass/Fail Criterion

Model	R	ΔR	R_C	ΔR_C	$R_{C'}$	$\Delta R_{C'}$
1. AFOQT-P, AFOQT-Q, FLYEXP, & UAV_NC	0.228**		0.325		0.470	
2. Model 1 + ATT_SR	0.232**	0.004	0.328	0.003	0.474	0.004
3. Model 1 + HTT_SR	0.235**	0.007	0.330	0.005	0.477	0.007
4. Model 1 + AHTT_SR	0.231**	0.003	0.327	0.002	0.473	0.003
5. Model 1 + AHTT_SR2	0.238**	0.010*	0.332	0.007	0.480	0.010
6. Model 1 + AHTT_SR3	0.236**	0.008*	0.331	0.006	0.479	0.009

N = 994

Notes. R values were based on the observed correlations. R_C values were based on the correlations corrected for range restriction. $R_{C'}$ values were based on the correlations corrected for range restriction and for dichotomization of the T-37 pass/fail score.

* $p < .05$, ** $p < .01$

As shown in Table 5, the observed, range restriction corrected, and fully corrected multiple correlations for the new baseline model (Model 1) versus the T-37 pass/fail criterion were 0.228, 0.375, and 0.470, respectively. An examination of Models 2 through 6 indicated that the amount of incremental validity observed by adding the ATT_SR, HTT_SR, AHTT_SR, AHTT_SR2, or AHTSR3 score to Model 1 was very small and ranged from 0.003 (Model 1 vs. Model 4) to 0.010 (Model 1 vs. Model 5). Only Model 5 (Model 1 + AHTT_SR2: R = 0.238) and Model 6 (Model 1 + AHTT_SR3; R = 0.236) demonstrated a statistically significant amount of incremental validity beyond the new baseline (Model 1). If Model 5 were adopted as the new PCSM model, it would be necessary to administer three TBAS subtests (ATT, HTT, and UAV). If Model 5 were adopted as the new PCSM model, it would be necessary to administer four TBAS subtests (ATT, HTT, AHTT, and UAV).

Table 6. Summary of Additional Regression Analyses for SUPT T-37 Total Score Criterion

Model	R	ΔR	R_C	ΔR_C
1. AFOQT-P, AFOQT-Q, FLYEXP, & UAV_NC	0.254**		0.360	
2. Model 1 + ATT_SR	0.260**	0.006	0.364	0.004
3. Model 1 + HTT_SR	0.255**	0.001	0.360	0.000
4. Model 1 + AHTT_SR	0.257**	0.002	0.362	0.002
5. Model 1 + AHTT_SR2	0.258**	0.004	0.363	0.003
6. Model 1 + AHTT_SR3	0.259**	0.005	0.363	0.003

N = 994

Notes. R values were based on the observed correlations. R_C values were based on the correlations corrected for range restriction.

* $p < .05$, ** $p < .01$

Similar regression analyses conducted using the T-37 Total Score criterion indicated that none of the additional TBAS scores (ATT_SR, HTT_SR, AHTT_SR, AHTT_SR2, or AHTSR3) incremented the new baseline model. See Table 6 for a summary of these analyses.

Additional Pre-Implementation Issues

Ree (2003b) identified several issues that should be addressed prior to operational implementation of TBAS in order to be in compliance with common test standards (American Psychological Association, 1999). Some of these were development of supporting documentation (test manual), subgroup analyses (subgroup norms, examination of test bias and adverse impact), and the development of test-retest norms and policy.

Work is currently in progress to re-host the TBAS on a new computer system that will administer the test battery via the internet (Strickland, 2004). Prior to operational implementation, it will be necessary to conduct an equating study to determine whether the tests administered on the preoperational system and the operational system measure the same psychological constructs, compare the score distributions of the two forms of the tests, and develop equating tables (Carretta & Ree, 1993). Equating is required so that scores from the operation version of TBAS can be used in pilot candidate selection regression equations developed on the basis of the preoperational form of TBAS.

CONCLUSION

A series of analyses were performed to evaluate the predictive validity of TBAS scores versus SUPT T-37 performance criteria and their incremental validity when used along with other measures of pilot training aptitude (i.e., AFOQT and previous flying experience). Although scores from several TBAS subtests showed predictive validity against T-37 performance, most of these failed to demonstrate incremental validity beyond a baseline pilot candidate selection model that included the AFOQT Pilot and AFOQT Quantitative composites and a measure of previous flying experience. Only the TBAS UAV number correct score demonstrated incremental validity over the baseline model versus both the T-37 pass/fail and the T-37 Total Score criteria. A small additional increment in validity was found for the T-37 pass/fail criterion for a TBAS number of skilled redirects composite score based on either ATT and HTT or on ATT, HTT, and ATTHTT.

Several additional issues were identified that should be addressed prior to operational implementation of TBAS and a revised PCSM model. These include development of documentation, subgroup analyses, development of test-retest norms and retest policy, and a study to equate the preoperational and operational forms of TBAS.

REFERENCES

Carretta, T. R. (2000). US Air Force pilot selection and training methods. *Aviation, Space, and Environmental Medicine, 71,* 950-956.

Carretta, T. R., & Ree, M. J. (1993). Basic Attributes Test (BAT): Psychometric equating of a computer-based test. *The International Journal of Aviation Psychology, 3,* 189-201.

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology, 8,* 29-42.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249-253.

Gould, R. B., & Shore, C. W. (2003). *Development of AFOQT Form S.* San Antonio, TX: Operational Technologies Corporation.

Johnson, J. T., & Ree, M. J. (1994). RANGEJ: A Pascal program to compute the multivariate correction for range restriction. *Educational and Psychological Measurement, 54,* 693-695.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62 (Pt.1),* 28-30.

Ree, M. J. (2003a). *Test of Basic Aviation Skills (TBAS) incremental validity beyond Air Force Officer Qualifying Test (AFOQT) pilot composite for predicting pilot criteria.* San Antonio, TX: Operational Technologies Corporation.

Ree, M. J. (2003b). *Test of Basic Aviation Skills (TBAS): Scoring the tests and compliance of the tests with the standards of the American Psychological Association.* San Antonio, TX: Operational Technologies Corporation.

Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology, 79,* 298-301.

Skinner, J., & Alley, W. E. (2002). *Air Force Officer Qualifying Test (AFOQT): Form R and S Development and Norms.* San Antonio, TX: Operational Technologies Corporation.

Strickland, T. (2004). *Air Force pilot selection assessments.* Paper presented at the 52nd meeting of the Department of Defense Human Factors Engineering Technical Advisory Group. Alexandria, VA.